Offer #2023-05689

Post-Doctoral Research Visit F/M Post-Doctoral Research Visit F/M Automatic speech recognition for non-natives speakers in a noisy environment

Contract type : Fixed-term contract

Level of qualifications required : PhD or equivalent

Fonction : Post-Doctoral Research Visit

Level of experience : Recently graduated

Context

This post-doc will be performed at MultiSpeech Team of INRIA-LORIA, Nancy.

MULTISPEECH is a joint research team between the Université of Lorraine, Inria, and CNRS. It is part of department D4 "Natural language and knowledge processing" of LORIA.

Its research focuses on speech processing, with particular emphasis to multisource (source separation, robust speech recognition), multilingual (computer assisted language learning), and multimodal aspects.

Assignment

Context
When a person has their hands busy performing a task like driving a car or piloting an airplane, voice is a fast and efficient way to achieve interaction. In aeronautical communications, the English language is most often compulsory. Unfortunately, a large part of the pilots are not native English and speak with an accent dependent on their native language and are therefore influenced by the pronunciation mechanisms of this language. Inside an aircraft cockpit, non-native voice of the pilots and the surrounding noises are the most difficult challenges to overcome in order to have efficient automatic speech recognition (ASR). The problems of non-native speech are numerous: incorrect or approximate pronunciations, errors of agreement in gender and number, use of non-existent words, missing articles, grammatically incorrect sentences, etc. The acoustic environment adds a disturbing component to the speech signal. Much of the success of speech recognition relies on the ability to take into account different accents and ambient noises into the models used by ARP.
Automatic speech recognition has made great progress thanks to the spectacular development of deep learning. In recent years, end-to-end automatic speech recognition, which directly optimizes the probability of the output character sequence based on the input acoustic characteristics, has made great progress [Chan et al., 2016; Baevski et al., 2020; Gulati, et al., 2020].

Objectives
The recruited person will have to develop methodologies and tools to obtain high-performance non-native automatic speech recognition in the aeronautical context and more specifically in a (noisy) aircraft cockpit.
This project will be based on an end-to-end automatic speech recognition system [Shi et al., 2021] using wav2vec 2.0 [Baevski et al., 2020]. This model is one of the most efficient of the current state of the art. This wav2vec 2.0 model enables self-supervised learning of representations from raw audio data (without transcription).

References
[Baevski et al., 2020] A. Baevski, H. Zhou, A. Mohamed, and M. Auli. Wav2vec 2.0: A framework for self-supervised learning of speech representations, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), 2020.
[Chan et al., 2016] W. Chan, N. Jaitly, Q. Le and O. Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 4960-4964, 2016.
[Chorowski et al., 2017] J. Chorowski, N. Jaitly. Towards better decoding and language model integration in sequence to sequence models. Interspeech, 2017.
[Houlsby et al., 2019] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, S. Gelly. Parameter-efficient transfer learning for NLP. International Conference on Machine Learning, PMLR, pp. 2790–2799, 2019.
[Gulati et al., 2020] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang. Conformer: Convolution-augmented transformer for speech recognition. Interspeech, 2020.
[Shi et al., 2021] X. Shi, F. Yu, Y. Lu, Y. Liang, Q. Feng, D. Wang, Y. Qian, and L. Xie. The accented english speech recognition challenge 2020: open datasets, tracks, baselines, results and methods. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6918–6922, 2021.

Main activities

The main activities are those typical of a postdoctoral intern.

They include: literature reading, scientific development, programming and simulation, data processing, reporting and presentation, paper writing, collaboration with the team, the supervisors and other scientific partners, participation to conferences and workshops.

Duration: 24 months

Skills

- Ph.D. degree in speech/audio processing, computer vision, machine learning, or in a related field,
- ability to work independently as well as in a team,
- solid programming skills (Python, PyTorch), and deep learning knowledge,
- good level of written and spoken English.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

2746€ gross/month

General Information

- **Theme/Domain :** Language, Speech and Audio
  IT Technical and production engineering (BAP E)
- **Town/city :** Villers lès Nancy
- **Inria Center :** Centre Inria de l'Université de Lorraine
- **Starting date :** 2023-03-01
- **Duration of contract :** 2 years
- **Deadline to apply :** 2023-06-30

Contacts

- **Inria Team :** MULTISPEECH
- **Recruiter :**
  Illina Irina / irina.illina@loria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated

in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.